

CAI
BSI
-1991
R34

Analytical Studies Branch

POHEM - A New Approach to the Estimation of Health Status Adjusted Life Expectancy

by

Michael C. Wolfson¹

No. 34

Research Paper Series



Statistics
Canada

Statistique
Canada

Canada



ANALYTICAL STUDIES BRANCH RESEARCH PAPER SERIES

The Analytical Studies Branch Research Paper Series provides for the circulation, on a pre-publication basis, of research conducted by Branch staff, visiting Fellows and academic associates. The Research Paper Series is intended to stimulate discussion on a variety of topics including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, statistical computing and simulation. Readers of the series are encouraged to contact the authors with comments, criticisms and suggestions. A list of titles appears inside the back cover of this paper.

Papers in the series are distributed to Statistics Canada Regional Offices, provincial statistical focal points, research institutes, and specialty libraries. Each paper is catalogued on the DOBIS computer reference system and in various Canadian university library reference systems.

To obtain a collection of abstracts of the papers in the series and/or copies of individual papers (in French or English), please contact:

Publications Review Committee
Analytical Studies Branch, Statistics Canada
24th Floor, R.H. Coats Building
Ottawa, Ontario, K1A 0T6
(613) 951-8213

CAI
BSI
-1991
R34

**POHEM - A New Approach to the Estimation of
Health Status Adjusted Life Expectancy**

by

Michael C. Wolfson¹

No. 34

Analytical Studies Branch
Statistics Canada
and
Canadian Institute for Advanced Research
1991

A paper presented to the first workshop of the
International Network on Health Expectancy
(REVES -- Réseau Espérance de Vie en Santé)
September 11-12, 1989
Quebec City

¹ I am indebted to Geoff Rowe, Jane Gentleman, Alice Nakamura, and an anonymous discussant for many valuable comments on earlier drafts. I of course remain responsible for the content.

The analysis presented in this paper is the responsibility of the author and does not necessarily represent the view or policies of Statistics Canada.

Aussi disponible en français.



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

<https://archive.org/details/31761116346842>

ABSTRACT

This paper describes a general methodology for estimating life expectancy adjusted for variations in health status during the course of individuals' lifetimes -- the population health model, POHEM. Measures such as "disability-free life expectancy" and the life table methodology on which these kinds of indices are based are considered. The restrictions embodied in such measures and their underlying methodologies can be conveniently avoided with the POHEM microsimulation approach. Prototypical outputs of POHEM are presented, and it is argued that the POHEM methodology is not unduly complex. Many countries could use it to generate health status adjusted life expectancy indices given already available data. Moreover, POHEM provides a framework for integrating a range of health data and for producing a family of important health indices.

Received: January 29, 1990

Accepted: January 17, 1991

Keywords: health status; microsimulation; longitudinal analysis;
life expectancy.

A. Introduction

This paper describes a general methodology for estimating life expectancy adjusted for variations in health status during the course of individuals' lifetimes -- the population health model (POHEM). This methodology is part of a broader project at Statistics Canada, the development of a new System of Health Statistics.

The context for POHEM, the System of Health Statistics, is briefly described in the following section. The paper then considers existing measures such as "disability-free life expectancy" and the life table methodology on which these kinds of indices are based. Such measures and the related methodologies are restrictive, so the paper turns to some of the major factors that health status adjusted life expectancy should take into account, and the methodology thereby required -- longitudinal microsimulation modelling. Some prototypical results using POHEM are presented, and the paper concludes by sketching several directions in which future development of POHEM could be extended.

B. The System of Health Statistics Project

Work is underway at Statistics Canada to develop a new conceptual framework for integrating health data -- a System of Health Statistics. This work stems from a broad view that there were basic difficulties with the current set of Canadian health-related statistical series. The initial review suggested that there were two major problems.

First, data collection efforts were seriously imbalanced, with relatively much more data collected on the inputs to the health care system, including financial costs, and much less on the supposed outputs of that system, namely the health status of the Canadian population.

Second, the various data collection efforts lacked coherence. This is a more abstract concern, but none the less important. Unlike other major statistical systems such as the System of National Accounts and population demography, health data do not "add up". They appear in compendia merely as a matter of juxtaposition; there are no underlying arithmetic identities or unifying theories.

In order to remedy these problems, a new conceptual framework for health statistics has been proposed, the System of Health Statistics as described in Wolfson (1989a). This proposed framework builds on three major existing intellectual foundations. The first, and most pertinent to this paper, is the extension of life table methodology to estimate various generalizations of life expectancy. The basic idea is to take account of how healthy people are during their lives, not just whether they are alive or dead. In turn, a summary index based on this notion would form a popular measure of overall health status -- a GDP (gross domestic product) or CPI (Consumer Price Index) of health.

The second foundation is input/output tables to be used to represent, in physical units, the variety and scope of activity involved in health care. The third foundation is the System of National Accounts, with its concepts of institutions, sectors, and factor costs.

In addition to these major intellectual foundations, the proposed System of Health Statistics conceptual framework adopts several developmental premises. One premise is not to limit the framework only to data that are currently collected. Instead, the operative constraint on the data to be included in the System of Health Statistics is that their collection should be technically feasible and not unduly expensive. This premise is necessary if the project is to provide useful guidance on data collection priorities, including new kinds of data.

This ability to consider new kinds of data is particularly relevant to POHEM, since it appears that more data on the longitudinal dynamics of health status should be collected. In this regard, major improvements are being undertaken in the administrative information systems underlying provincial health care programs. As a result, the marginal costs of a range of individual longitudinal health-related microdata can be expected to drop significantly.

The second major developmental premise is to recognize the power of modern computing. The revolutionary decline in the costs of computation opens major possibilities that simply were infeasible at the inception of the National Accounts, for example. It is no longer necessary to be restricted to aggregate totals or partially disaggregated sub-totals. Thus, the System of Health Statistics is presumed to be built on an explicit microdata foundation. In turn, the System of Health Statistics is envisaged as being much more than a compendium of printed numeric tables. Its full realization should be a database combined with retrieval and analytical software -- implemented in a portable and broadly accessible software environment.

Figure 1 gives an overview of the main elements of the System of Health Statistics and the key linkages among them.² There are two main arrays of data that are fundamental. The upper three-dimensional array represents the population. The lower three-dimensional array represents the resources devoted to health care.

These two arrays are connected by a flow of "visits" whereby people utilize health care services, for example by receiving treatments. In turn, these "visits" or treatments may be efficacious, and thus affect peoples' health. To the right, three main groups of summary statistics are shown which can be derived from these two arrays of data.

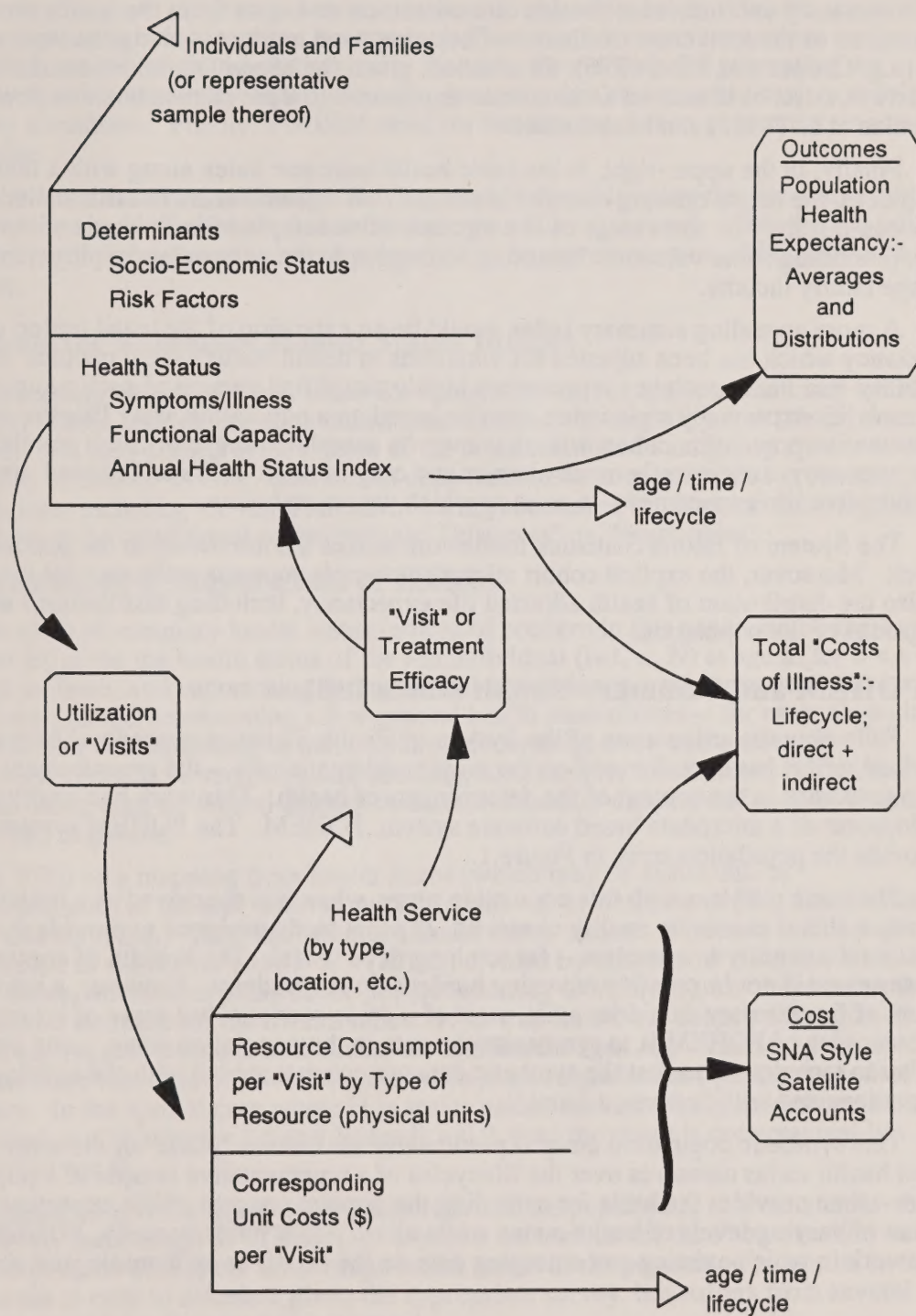
To give a bit more detail, the upper population array represents a dynamic view of a population cohort with advancing age through its lifecycle. The vertical axis represents various attributes of a given individual at each age (his or her "state space") -- socio-economic status (e.g. income, marital status), risk factors (e.g. blood pressure, smoking, work-place stress), symptoms and illnesses (e.g. coronary heart disease), functional capacity (e.g. gross motor, vision), and an overall annual health status index based on some combination of the just-mentioned variables. The third axis (going back into the page) represents various individuals in the population, and includes their familial structures.

The lower health care resource array has the same horizontal axis -- which is best thought of as indicating age-specific differences in health care services, other things being equal. The top half of each column vector in the array then represents consumption in physical units of specific resources -- such as operating room hours, bed-days, and physician office visit hours -- associated with a "visit" to a particular health care provider for a given malady. The bottom half of the vector gives their corresponding unit costs. The third dimension represents a combination of attributes of the health care "visit" -- the kind of malady, the kind of treatment, the type of health care institution, and its geographic location. (Subscripts are inexpensive in theory!)

These two basic arrays of data can be used to compute three major groups of statistics or social indicators. In the lower left is the conventional bottom line -- aggregate dollar costs. As

2 This diagrammatic representation, from Wolfson (1989a), can be criticised on several grounds. In particular, it appears unduly reductionist, and gives virtually no weight to environmental determinants of health -- psycho-social as well as physical. These concerns are being addressed in Wolfson and Glynn (1990).

The System of Health Statistics Framework



well, the physical resource and unit cost data in the lower array provide the basis for a "satellite account" for the health sector in the conventional System of National Accounts manner (e.g. Pommier, 1981, Teillet, 1988).

In the middle left, drawing on both individual lifecycle disease histories from the population array and individual health care utilization and costs from the health care array, are calculations of the total costs of illness -- both direct and indirect -- along the lines pioneered by Rice (e.g. Cooper and Rice, 1976). In addition, given the lifecycle orientation of the two basic data arrays, costs of illness on an incidence as opposed to a prevalence basis as developed by Hartunian et al. (1981) can be estimated.

Finally, in the upper-right, is the basic health outcome index along with a family of sub-indices, the major missing element in virtually all national sets of health statistics. This index could simply be an average over a representative sample of individuals of their health status (assuming some valuation function), analogous to the national unemployment rate or average family income.

A more appealing summary index would be an extension of the usual notion of life expectancy which has been adjusted for variations in health status during peoples' lifetimes. Disability-free life expectancy represents a highly simplified version of such a summary index. Any such life expectancy style index must be based on a population array like the one shown -- a representative population cohort with data over its complete lifecycle. Such population arrays, while necessary, are typically much simpler and only implicit in life table based measures like disability-free life expectancy -- a point to which we return below.

The System of Health Statistics framework makes the microdata in the population array explicit. Moreover, the explicit cohort microdata sample supports estimates not just of averages, but also the distribution of health adjusted life expectancy, including distributions among various sub-groups of the population.

C. POHEM and Summary Health Status Indices

Following the articulation of the System of Health Statistics conceptual framework just described, effort has been focused on the most problematic part -- the measurement of population health outcomes in the context of the determinants of health. This work has involved the development of a microdata based software system, POHEM. The POHEM system is designed to provide the population array in Figure 1.

The basic problem with this population array, when it is conceived as a longitudinal cohort data set, is that it cannot be readily observed. A panel study designed to provide the data would take almost a century to complete -- far too long to be useful. The breadth of content would also impose severe if not impossible reporting burdens on respondents. However, a substantial amount of fragmentary data does exist, most of it cross-sectional but some of it longitudinal. The basic idea of POHEM is to synthesize the data in the population array, using computer simulation techniques, so that the synthetic data are consistent both with the existing fragmentary observations and with "common sense".

This synthetic population array -- particularly the bottom "plane" of the array giving annual health status measures over the lifecycles of a representative sample of a population cohort -- then provides the basis for extending the popular concept of life expectancy to take account of varying levels of health status while alive. More fundamentally, POHEM provides a framework in which existing and emerging data on the causal or epidemiological linkages among

risk factors, disease precursors, diseases, functional limitations, and a summary index of health status can all be incorporated. (Similar objectives have been expressed by others, for example Caselli et al., 1989.)

Of course, POHEM is potentially very complex, so is it really necessary to develop such an elaborate methodology in order to generate a reasonable summary index of health status? Our affirmative response is based on three main points. First, current summary indices embody gross simplifications that render them unsatisfactory. Second, the methods for generating the population arrays that implicitly underlie existing summary health status indices are restrictive and seriously unrealistic. Finally, POHEM need not be that complex to generate a reasonable summary index.

We shall start in the immediately following section by addressing the first two of these points. Current summary indices and their limitations will be discussed. Then the last point will be established in the subsequent section by describing the current version of POHEM and some of its outputs.

D. The Basis for Summary Health Status Indices

In this section, we develop a notation for describing various kinds of summary health status indices. These include life table-based measures such as disability-free life expectancy as special cases. We then discuss briefly some of the considerations in assigning a value to different states of health. Finally, this section considers the various methods for constructing summary indices, including the implications of taking account of information showing health state transitions to be conditional on individuals' "histories" or "biographies".

1. Representing a Simple Class of Indices

The class of summary health status indices of concern in this paper can be represented as sums. Let $H(i,a)$ be the health status of the i -th individual ($i=1, \dots, N$) at age a (for $0 < a < A_i = i$'s the age at death, and suppressing for notational convenience a subscript denoting gender). H could be an integer representing a few general health state classes -- for example an integer from one to four corresponding to non-, mildly, moderately, or severely disabled. Alternatively, H could be vector-valued showing, for example, five or six levels of functional limitations along each of several distinct disease or disability dimensions such as vision, mobility, and cognition.

Let $V(H)$ be a mapping from health status (which may be either uni- or multi-dimensional) to the unit interval, with $V(\text{dead}) = 0$ by definition. (We consider later the basis for this mapping.) Then the summary health status indices of interest in this paper can be represented as sums over i and a of $V(H(i,a))$ divided by the number of individuals, N . (In Figure 1 above, the bottom plane of the population array is $V(H(i,a))$, i.e. the matrix of elements to be summed for the overall index. Note that for a birth cohort, the right side of this array will be "ragged" because individuals die at different ages A_i .) When the i 's index individuals from the same birth cohort, this index is a straightforward generalization of life expectancy. In the special case where H is scalar valued and takes only one of two states, alive or dead, and $V(\text{alive}) = 1.0$ and $V(\text{dead}) = 0.0$, then the result is conventional life expectancy at age 0.

Alternatively, if the $H(i,a)$ are representative of the population, as from a cross-sectional health survey, the average of $V(H(i))$ (age a is now an attribute of the i^{th} observation and no longer independent of i) gives the average health status of the population at a point in time. This measure is easy to calculate given the appropriate survey, but suffers from several limitations, as will be discussed in a moment.

The basic problem with life expectancy as a summary indicator of health status is that it takes no account of how healthy individuals are when they are alive. Two societies could have identical mortality rates, and hence life expectancies, but still be widely different in the prevalence of chronic diseases and functional impairments. Life expectancy is not sensitive to variations in morbidity.

One group of measures that attempts to address this concern is "disability-free life expectancy" or DFLE (e.g. Robine, 1986) -- the number of years individuals can expect to live free of disability. This measure can be represented in our notation by a variable $H(i,a)$ that takes one of two values -- alive and well, or alive and disabled -- and a V function that assigns values of 1.0 and 0.0 respectively to these states. Given these specific $V()$ and $H()$ functions, DFLE is then the sum over i and a of $V(H(i,a))$ divided by N . Thus, disability-free life expectancy treats the disabled as if they were dead.

This is clearly a very simplified approach to taking account of health status. It is insensitive to all changes in morbidity except those that cause an individual to cross the borderline between "alive and well" and "alive and disabled". A similar criticism applies to related measures like "non-institutionalized life expectancy" (Crimmins et al., 1989) and "active life expectancy" (Katz et al., 1983).

A richer approach is that used by Wilkins and Adams (1983). $H(i,a)$ in their analysis could take any one of five states valued as follows: $V(\text{no disability}) = 1.0$, $V(\text{short-term impairment}) = 0.5$, $V(\text{chronic minor restriction}) = 0.7$, $V(\text{chronic major restriction}) = 0.6$, $V(\text{chronic severe disability}) = 0.5$, and $V(\text{institutionalized}) = 0.4$, with $V(\text{dead}) = 0.0$. However, these values are admittedly approximate, and the range of states is still quite coarse.

2. Valuing Health States

Life expectancy, as just described, treats individuals as being in one of two distinct states -- alive or dead, while the Wilkins and Adams (1983) estimates consider six distinct states. The logical extension of the concern about the "sensitivity" of the summary health index to finer gradations in health status is to consider individuals as being in any of a continuum of states ranging from alive and in full health, through various degrees of illness or functional limitation, to dead, and even to states of being alive that are considered worse than death -- such as very painful terminal cancer.

By convention, full health is given a value of 1.0 and death a value of 0.0, with various morbid states given a value between zero and one, and rarely a value less than zero. (Some ecstatic states might be given values greater than 1.0.)

For any given person-year of life, assigning a value to health status involves two main steps. The first is to classify the person's health status in some descriptive way; the second is to assign a value to that specific health state. For both of these steps, there are a variety of methods and approaches in the literature.

In the case of health states, the range of approaches includes clinical descriptions often based on the International Classification of Diseases, self-report descriptions of illnesses and health problems, and self- or other reports of functional limitations. Arguably, the latter approach is most appropriate. What matters most to an individual is whether he or she is in pain, or is able to get around the house, rather than the particular clinical name for the malady that gives rise to these health problems and functional limitations.

Data on the distribution of the population among various states of functional limitation are generally available, for example in the 1990 Ontario Health Survey (Statistics Canada and the Ontario Ministry of Health, 1989) and the 1986 Canadian post-censal Health and Activity Limitations Survey (Statistics Canada, 1988). The Health and Activity Limitations Survey includes a set of twenty functional limitation questions.

Similar sets of questions ("activities of daily living" and "instrumental activities of daily living") are available from population surveys in several countries, for example the U.S. National Health Interview and related Surveys (Erickson et al., 1988), the U.S. Long Term Care longitudinal survey (Manton and Stallard, 1990), and the U.K.'s Office of Population Census and Surveys' disability surveys (Bebbington, 1989).

Turning to valuation, the 1990 Ontario Health Survey is particularly interesting because it includes a series of specially structured questions explicitly designed as the basis for health status valuation. These questions elicit health status in a number of discrete levels along eight different dimensions of functional health: hearing, seeing, communicating, gross motor, dexterity, cognitive, emotional, and pain. In order to assign values to each of these eight-dimensional health states, plans are underway to estimate a population-based multi-attribute value scale or utility function along the lines developed by Torrance (1987). Similar valuation methodologies have been developed by Rosser and Kind (1978), and Kaplan and Bush (1982), though critical comments are contained in Loomes and McKenzie (1989).

The methodologies to be described for estimating health status adjusted life expectancy will thus assume that a multi-attribute mapping has been estimated that can take functional health states into a scale or value in the $[0,1]$ interval -- the $V(\cdot)$ function in our notation.

3. Estimating the Distribution of the Population Among Health States

a. Population Averages or Cohort Expectancies

Given a health state valuation function, the next question is the source of the estimated distribution of these health states among the population, the $H(i,a)$ array. One obvious possibility is a cross-sectional survey like the 1986 Canadian Health and Activity Limitations Survey or the 1990 Ontario Health Survey. The basic formula given above will in this case simply yield the average health status in the population. However, these health status observations are disjoint. They do not represent individual lifecycle paths and cannot yield expectancies, just as the average age of the population in a given year is not the same as its life expectancy.

Why is a life table expectancy style measure generally preferable to a population average? First, there would be major interest in the trend over time in such a measure. However, trends in the population average will be confounded by changes in the age structure that have nothing to do with mortality (e.g. declines in fertility rates leading to an aging population, hence a larger proportion with low average health status). This confounding is avoided by using a stationary age structure as is implicit in life table style methods. Second, this population average takes no account of mortality, so a catastrophe or epidemic that killed all individuals with low health status would result in an increase in average health status -- among those surviving. Health status adjusted measures of life expectancy allow both mortality and morbidity to be combined in a single index. For these and other reasons, we shall pursue measures that build on the concept of cohort life expectancy.

Of course, life expectancy style measures are hypothetical; they do not refer to a real population as does an average. They are, on the other hand, based on current rates of flow or transition among states (e.g. recent mortality rates by age and sex). They are thus "purer" -- in the sense that they are not an accumulated melange of previous mortality rates of various dates applied to different birth cohorts as is the case with averages over current cross-sectional surveys of the population, like average age. Life expectancy (on a period basis) is akin to velocity rather than position, or to the first derivative of our current position, or the slope of society's current trajectory. Thus it gives (a linear approximation to) an indication of where society is headed.³

b. Cross-Sectional Disability Prevalences

With these arguments for measures based on life expectancy, a key issue is how the underlying $H(i,a)$ array is to be derived. In the case of disability-free life expectancy and the Wilkins and Adams measures, a straightforward mixture of lifetables and cross-sectional prevalences are used. The underlying individual life paths are never explicit -- i.e. the $H(i,a)$ array is never constructed, only the margins of this array are constructed. A period life table is used to derive a survival curve $S(a)$ giving the proportion of individuals expected to be alive at age a , given age- (and sex-) specific mortality rates for a recent year (or more typically the proportion of a birth cohort's potential life-years actually lived in the interval from age a to $a+1$).

A cross-sectional survey (or other equivalent data source) is then used to estimate a matrix of disability prevalences $P(a,d)$ where each row of the matrix (assuming a is discrete) gives the distribution of the (living) population of that age among the various disability states (possibly multi-dimensional $d = 1, \dots, D$). In the simplest case ($D = 1$), $P(a,d) = P(a) =$ "the" percentage disabled. There are no distinct individuals, so that the average of $H(i,a)$ over i (including those individuals who have already died, i.e. $a > A_i$), denoted $H(.a)$, is simply $S(a) P(a,d)$. In other words, the percentage of the original birth cohort in the d -th health state at age a is the product of a lifetable calculation of the proportion of a hypothetical birth cohort expected to survive to age a , times the observed cross-sectional fraction of (living) individuals age a in that health state. (For convenience, we continue to ignore the notational differences between lifetables with discrete ages, and survival curves based on continuous age.)

This method of deriving disability-free life expectancy makes it clear that such indices are more akin to age-standardized mortality rates than to life expectancies. Essentially, a period lifetable population is being used to estimate an age distribution weighted sum of age-specific disability prevalences. Then life expectancy is multiplied by one minus the (age-standardized) average prevalence of disability to obtain disability-free life expectancy (assuming the zero/one health state value function described earlier).

3 Of course, period life expectancy estimates embody history insofar as the mortality rates being measured "today" for a given age group apply only to those who have survived to "today". This population of survivors is likely different from the population that would have survived had "today's" mortality rates applied in all previous years.

Disability-free life expectancies estimated in this way are "neither fish nor fowl" -- they are based on a mixture of stocks (age-specific disability prevalences) and flows (age-specific mortality rates). A purer or more consistent conceptual framework would use not only current mortality rates, but also current transition probabilities among disability states. Moreover, disability state and mortality transitions are not independent. Nor are transitions among disability states unidirectional; many people who become disabled in some way subsequently recover.

c. Multi-State Life Tables

While data on such disability state transitions are more difficult to obtain, the resulting estimates of disability-free life expectancy appear to be sensitive to the use of such data. For example, Brouard and Robine (1989) have shown that the mortality rate of the institutionalized population is significantly higher than that for the non-institutionalized, and taking this into account has an important effect on estimates of disability-free life expectancy. Similarly, Branch (1989) shows, drawing on longitudinal survey data, that taking account of the fact that health status does not always get progressively worse but often improves over the life course has a material impact on the estimates.

In order to use these kinds of transition data in disability-free life expectancy estimates, multi-state life table techniques are used. Instead of representing only one transition -- from alive to dead -- as in conventional life tables, transitions among several living states and death are explicitly considered. In Branch (1989), for example, the life table has three states: alive and active, alive and inactive, and dead. Thus, in our notation, the values in the a -th row of the $H(a)$ matrix (i.e. age by health state) are the product of the values in the previous row and a 3×3 transition matrix. In other words, the values in the a^{th} row are calculated as a linear function of the $a-1^{\text{st}}$ row in the H matrix.

d. Semi-Markov Processes

A further extension of the methodology is illustrated in Loewy et al. (1989) in the context of evaluating outcomes of clinical drug trials. In their case, four states of health are considered based jointly on history of heart attacks and complications of the drug therapy being assessed. More importantly, data on the duration dependence of the transition probabilities are taken into account. The process describing the probability of heart attack estimated from the longitudinal follow-up in the drug trial depends, among other things, on how long it has been since the individual had a previous heart attack.

This dependence, empirically, is not first order Markov as is conventionally assumed in single and multi-state life tables, so a semi-Markov representation is used. In terms of the $H(a)$ matrix, each row is calculated according to a more complicated process, potentially a non-linear function of all the rows at earlier ages, not just the immediately preceding row.

However, even this representation is still a gross simplification. For example, Wolfson et al. (1990) show that mortality is significantly associated with marital status, and with income -- including latent effects over periods as long as decades. As another example, the progression of senile dementia depends on whether it is Alzheimers or multi-infarct (Forbes and Barham, 1989).

It should be clear from these examples that transitions among health states follow complex patterns. Each transition probability is much better thought of as a function of an individual's multivariate life path up to the given age, not simply a scalar. Failure to take account of this complexity can be expected to result in poor and possibly misleading estimates of $H(a)$, and thus of health status adjusted life expectancy. In essence, it is important to relax at least some of the assumed homogeneity embodied in the usual life table-based estimates of disability-free life expectancy. (An example of the serious biases that result from failure to take account of these more complex types of hazards is given in the case of marriage breakdown in Rowe and Wolfson, 1990.)

e. Microsimulation

But is it reasonable to extend life table techniques to meet these concerns? Generally, the answer is no; it would not be practical. To show this, we can consider the following relatively straightforward extension of current life table methods.

Individuals are classified into four states of functional limitation -- healthy, mildly limited, moderately limited, or severely limited. For each age (within a five year range, say) and sex, there is a four by five matrix of transition probabilities from one of these four health states to the same four health states or death.

But because the transition probabilities vary with duration, this matrix only applies to individuals who were in the last health state for less than five years. Another matrix gives corresponding transition probabilities for individuals who have been in their previous health state from five to ten years, another matrix for durations of ten to fifteen years, and so on to a matrix for those who have been limited since birth.

This situation means that an individual of a given age and sex can be in one of up to 40 or more different health states -- four categories of functional limitation times ten (or more or less, depending on age) duration categories.

Furthermore, mortality rates depend on whether or not the individual is institutionalized, and whether or not the individual is married (recall Brouard and Robine, 1989, and Wolfson et al., 1990). Even if mortality is assumed to be independent of functional limitation, to keep the population partitioned into exhaustive and mutually exclusive categories, the 40 health states have to be each broken down by marital status and institutionalization -- yielding 160 different health status/demographic states.

At this point, it can be seen that an abridged (i.e. five yearly rather than annual) life table approach would require a multi-state life table with 160 columns. This is manageable, but unwieldy. Furthermore, it would not be very flexible. Suppose, for example, that the transition from married to single were to be represented more carefully using divorce rates and mortality rates for the spouse as well as remarriage rates, as in a marital status life table (Adams and Nagnur, 1988). Then four rather than two marital status categories would be required (single, married, widow(er)ed, divorced), bringing the number of columns required in the multi-state life table up to 320. The change would also require recalculating all the age/sex specific transition probabilities.

Finally, if hazard analysis based on longitudinal data were available, then the number of required columns would explode. For example, Rowe (1989) shows that the risk of marriage breakdown (divorce or separation) depends in a complex way on the duration of the marriage, and for females whether they were teenage brides. The epidemiological literature, of which the Alameda County Study (Berkman and Breslow, 1983) is a well known example, is replete with estimates of such hazards.

Note that all of the extensions to the life table just mentioned are based on empirical results which are already well known, and available in a suitably quantified form. However, incorporating such empirical results in a multi-state life table framework is clearly impractical. Vaupel and Yashin (1986) conclude -- for a related set of models -- that the only practical solution when more realistic state space and transition probability functions are used is microsimulation. (This is by no means a new argument; see Nakamura and Nakamura, 1978.)

E. The POHEM Methodology

Fortunately, there is an alternative methodology which can draw on such complex multivariate transition probability data, and which in the limit is computationally equivalent. This is monte carlo microsimulation.

The idea of monte carlo microsimulation in this case is to synthesize a sample of complete individual life histories. Exactly the same transition probabilities are used as would be used in a multi-state life table analysis (or semi-Markov process, for that matter). In the simplest example of a single state life table, the synthesis of each individual proceeds by "creating" that individual at birth. A random number is then drawn in the unit interval to see whether that individual dies in his (or her) first year -- by checking to see whether the number drawn is less than the mortality rate for that age. If so, the individual "dies". Otherwise, he survives to the next year where the process is repeated, and so on.

This process is completed 100,000 times, say, and for each individual the result is an age at death. Then from this sample, an average age at death is easily computed. As the synthetically generated sample of these ultimately simple life histories is increased in size, it approaches exactly the result that would be obtained from the life table calculation.

This monte carlo process is clearly far more computationally intensive than a life table for deriving simple life expectancy. But now suppose that each individual is exposed to multiple "hazards" -- for example becoming married, becoming moderately or severely disabled, as well as dying. In addition, each hazard depends in a complex multivariate way on previous states. This is the current situation, as outlined in the previous section. Then it is straightforward to write an algorithm for each kind of transition that computes an individual's specific probability of making that transition in that year conditional on his or her current state and previous "biography" or life path through various states.

This is precisely what POHEM, and the Cohort Lifecycle Simulation System of which it is a part, do. They incorporate specific algorithmic realizations of estimated transition probability functions for a variety of states.

For example, DEMOGEN is the demographic companion model to POHEM. It incorporates all the elements of marital status life tables such as in Adams and Nagnur (1988), plus labour market activity. Table 1 lists the processes included in DEMOGEN and the variables on which the transitions are conditional. Analyses of longitudinal demographic data from the

1984 Family History Survey by Rowe (1989) have provided union formation and dissolution probabilities, and a similar analysis by Picot (1989) underlies the labour force transitions. Two applications of DEMOGEN are described in Wolfson (1989b).⁴

Table 1 Processes Represented in DEMOGEN

Process or Event	Conditional Upon
Educational Attainment	"pre-ordained" (endowed at birth by drawing from observed joint distribution of husband - wife educational attainments)
First union (legal marriage or common-law union - CLU)	age, sex, education, fertility (for females), labour force experience, previous CLU, pre-ordained "marriageability"
First Spousal Age Difference	age at marriage, observed distribution
Fertility	age, parity, marital status
Child Custody	marital status
Child Separation (leaving home)	age, sex, birth order of child
Union Dissolution (divorce or separation)	age, duration of marriage, presence of children, labour force experience, age at marriage
Remarriage	age, sex, divorce versus widow(er)
Second Spousal Age Difference	marrying person's age at marriage, sex, prior marital status, observed distributions
Labour Force Participation Year by Year	age, sex, marital status, presence of children by age group, educational attainment, duration in state
Labour Market Potential Earnings	age, sex, labour force attachment
Mortality	age, sex, marital status (unless used in conjunction with POHEM)

It is relatively straightforward, in this software framework, to incorporate complex multivariate transition probability functions among health states. POHEM thus builds on the socio-demographic structure provided by DEMOGEN. This strong connection between DEMOGEN and POHEM is not only a matter of convenience. It is clearly the case that morbidity and mortality are strongly correlated with socio-economic status -- for example marital status, income, occupation, educational attainment, and social support -- even after controlling for conventional bio-medical risk factors like hypertension, smoking, and obesity (e.g. Marmot, 1986; Berkman and Breslow, 1983).

4 Note that DEMOGEN and POHEM are discrete time dynamic microsimulation models. A number of alternative simulation strategies have been developed. For example Hayes (1989) suggests that a continuous time, discrete event formulation may be superior for these kinds of problems. However, this "event history" approach to microsimulation modeling would be problematic in our context. For example, DEMOGEN modules like those for union formation and dissolution incorporate competing risks and time varying covariates, and are based on longitudinal data sets having discrete time steps. The discrete event approach suggested by Hayes would not be suited to these aspects.

POHEM starts this process of building on the socio-demographic sub-strate of DEMOGEN with a module for the longitudinal progression of risk factors. Using data from the 1978 Canada Health Survey, the most recent comprehensive survey in Canada, transition probabilities have been estimated for the four main Framingham study coronary heart disease risk factors -- obesity, smoking, blood pressure, and cholesterol. A new methodology has been used to estimate "reasonable" transition probabilities in the absence of longitudinal data (Gentleman and Robertson, 1989). These transition probabilities are constrained so that they will reproduce the observed quadrivariate joint distribution of the risk factors within each sex across adjacent five year age groups. (The methodology uses optimization and a special form of linear programming -- network flow.)

In turn, these risk factors are used as inputs to a Coronary Heart Disease (CHD) module. It is based on the same underlying data as in Weinstein et al. (1987). The basic transitions in this CHD module are illustrated in Figure 2. The process starts by assessing for each year and for each individual with no prior CHD history his or her risk of a CHD event -- based on a risk function estimated from the Framingham Heart Study. If there is a CHD event, it is determined (according to an observed probability distribution disaggregated by age and sex) to be either a myocardial infarction (MI), a cardiac arrest (CA), both MI and CA, or angina. Subsequent CHD progression then follows the remaining paths in Figure 2.

Similarly, senile dementia onset and progression transition probabilities have been incorporated based on the analysis of Forbes and Barham (1989), and breast cancer based on work by Hill (1989). (More recently, a module describing the onset and progression of lung cancer, as a function of smoking and radon exposure, has been added, Gentleman et al, 1990.) Finally, a rather arbitrary set of values have been assigned to the various disease states -- on a provisional basis in order to illustrate the POHEM system.

F. Prototypical Results

Table 2 shows the basic results for the first three disease modules for CHD, dementia, and breast cancer onset and progression. These results show, for example, that just over half of all females (54.3%⁵) can expect at least one CHD event in their lifetimes -- based on the version of the Framingham CHD risk function used in Weinstein et al (1987), risk factor prevalences as observed in the 1978 Canada Health Survey, and the competing risks of other cause mortality based on 1986 Canadian cause-, age- and marital status-specific mortality rates.⁶ These CHD events occur on average at age 71.9, with not quite half (46.3%) associated with death from this cause. CHD deaths occur on average at age 78.3, one year later than life expectancy at age zero. This is not surprising, however, because CHD has a late onset, and those who die of it must have first survived long enough for the disease to set in.

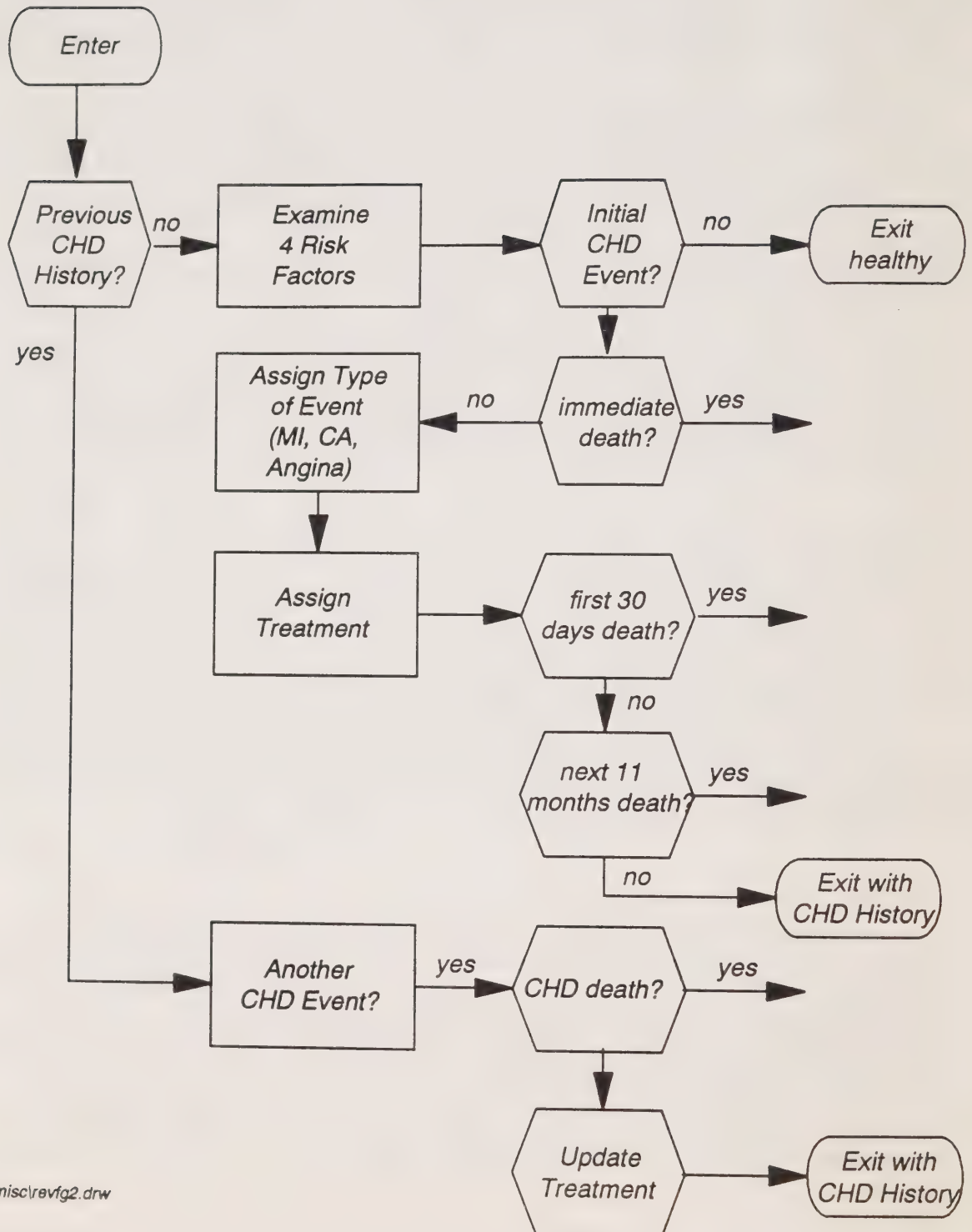
5 We use the exact number in the table in order to aid the reader in following the discussion. However, monte carlo sampling error results in estimates like this one only being precise to within several percentage points. Similarly, the life expectancy and average age at onset figures should only be considered accurate to within about one-half a year. Rowe and Wolfson (1989) illustrate the sample re-use method of variance estimation that has been incorporated into DEMOGEN and POHEM which underlies the preceding comment on the precision of the estimates.

6 In turn, marital status depends in a complex way on prior fertility and labour force history, educational attainment, and the survival of the spouse, as modelled by DEMOGEN.

POHEM - Coronary Heart Disease (CHD)

Module Based on Weinstein et al. (1987 AJPH)

(Annual Time Step)



Dementia, based on the analysis of incidence and progression provided by Forbes and Barnham (1989), has a much later average age at onset and is considerably less prevalent in a lifecycle sense -- 10.3% of females and 5.4% of males can expect ever to have dementia. Due to a lack of data, the dementia module assumes that dementia is never fatal. Breast cancer, on the other hand, has about an 85% case fatality rate (8.6% of women out of a total of 10.5% having incident cases). These fatalities occur on average about three years after onset, based on the detailed incidence and progression data provided by Hill (1989).

Recalling our earlier discussion of disability-free life expectancy type measures, these figures in Table 2 show that quite a wide range of statistics other than the usual expectancies can be computed using POHEM/DEMOGEN style microsimulation methods. They include the proportions of the population to whom a given event or set of events occurs, and the average waiting time until the set of events.

Table 2: Illustrative POHEM Output -- Average Ages and Percentages Involved for Key Health Events

Event	Females		Males	
	Age	Percentage*	Age	Percentage*
First CHD Event	71.9	54.3	64.5	56.5
Second CHD Event	77.1	36.2	70.8	48.5
Third+ CHD Event	80.5	42.4	74.9	43.6
CHD Death	78.3	46.3	72.9	55.8
CHD Death	78.3	25.2	72.9	31.5
Dementia Onset	79.8	10.3	78.7	5.4
Breast Cancer Onset	63.5	10.5	0.0	0.0
Breast Cancer Death	66.3	8.6	0.0	0.0
Other Cause Death	78.3	66.2	72.1	68.5
All Death (= life expectancy)	77.3	100.0	72.3	100.0
Population Health Expectancy (PHE)	74.5	100.0	69.9	100.0

* Percentage of all individuals in the cohort experiencing the event. Note, however, that the denominator for the percentages in the cases of "Second" or "Third+ CHD Event" are of those with a first or second CHD event respectively; and the denominator in the first of the two rows labelled "CHD Death" is those having had at least one CHD event.

Source: POHEM simulation run, sample size 5,000.

In addition to average ages and proportions involved in various disease events, and most importantly for the purposes of this paper, the "bottom line" of Table 2 shows estimated health status adjusted life expectancy.

A variety of names have been used in the literature for this measure, for example disability-free and active life expectancy as noted earlier. QALY (quality adjusted life year) is often used to describe the result of applying the $V()$ function (in our notation) to a given life year, while Kaplan and Bush refer to QWB (quality of well being), and Mehrez and Gafni (1989) use HYE (healthy-years equivalent) for the same purpose.

Notwithstanding the acronymic clutter entailed, we propose the term Population Health Expectancy, or PHE. This serves to distinguish the measure produced by POHEM or equivalent methodologies from other efforts for two basic reasons. First, PHE embodies a continuum of health states at each year of age rather than a simple dichotomy like disability-free or active life expectancy. Second, PHE is based on a representative and heterogeneous sample of (hopefully) realistic individual life paths, not an aggregation of life years from a set of discrete age and health state categories from a multi-state life table.⁷ We keep the preface "population" rather than just using "health expectancy" to emphasize the underlying heterogeneity of the population that is being taken into account.

The results shown are still primarily for illustrative purposes, and various forms of validation are underway. Nevertheless, they appear plausible. For both males and females, PHE is about three years less than unadjusted life expectancy.

Since these estimates are based on an explicit sample, it is also straightforward to calculate PHEs over shorter intervals like decades, and then to tabulate the distribution of these PHEs. Since some individuals will die during any age interval, PHEs within these intervals are computed as total "weighted" person-years lived (weighted by the $V()$ function) divided by the total number of unweighted person-years. Indeed, these PHE estimates can also be disaggregated by the value of the weighted person-years, for example between 80 and 89 percent. This is shown in Table 3. None of the three diseases being modelled are assumed to have onsets before age 25, so that everyone age 24 or less has a PHE of 100 percent. For the three diseases explicitly modelled, males generally have lower average PHEs within age ranges than females. The table clearly shows the increasing dispersion of PHE with advancing age.⁸

7 The life paths generated by POHEM are explicit, albeit synthetic. They are realistic in so far as they are constrained to aggregate to the various "marginal" data sets used to generate them. Moreover, micro- as well as macro-realism can be assessed by "browsing" through the sample of synthetic life paths. In contrast, life paths are not explicit in multi-state life tables. However, there is always an implicit set of synthetic life paths underlying any multi-state life table. Because these life paths are only implicit, they are never subject to scrutiny, so it has not been a concern that a portion are necessarily utterly unrealistic. Both represent a single birth cohort rather than a cross-section of a population of overlapping birth cohorts.

8 The bimodality of the distribution of PHE at the highest ages is due to our assumption, noted earlier, that dementia is never fatal. After progressive deterioration, individuals with dementia are assumed to continue life with $V(\text{dementia of long duration}) = 0.1$

Table 3: Illustrative POHEM Output -- Distributions of PHE Within Age and Sex Groups

PHE Levels (%)	Age of Female								
	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
0- 9	0.0	0.0	0.0	0.0	0.0	0.0	0.5	4.0	15.7
10-19	0.0	0.0	0.0	0.0	0.0	0.1	0.9	3.3	2.7
20-29	0.0	0.0	0.0	0.0	0.0	0.1	0.7	2.1	1.6
30-39	0.0	0.0	0.0	0.0	0.0	0.2	0.9	2.4	1.3
40-49	0.0	0.0	0.0	0.0	0.2	0.3	1.1	2.7	2.7
50-59	0.0	0.0	0.1	0.3	0.3	0.8	2.5	4.1	2.9
60-69	0.0	0.0	0.0	0.2	0.7	2.3	5.2	8.4	5.1
70-79	0.0	0.0	0.5	1.7	4.5	9.6	19.4	30.4	41.9
80-89	0.0	0.0	0.9	2.8	5.3	8.3	15.4	11.9	8.3
90-99	0.0	34.1	34.1	32.8	31.9	30.9	24.2	12.7	6.7
100	100.0	65.9	64.3	62.1	57.0	47.4	29.2	18.0	11.2
Cohort Surviving (per 1,000)	999	995	984	947	872	738	477	193	45
Average PHE	1.000	0.983	0.980	0.974	0.962	0.932	0.852	0.729	0.631

PHE Levels (%)	Age of Male								
	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
0- 9	0.0	0.0	0.0	0.0	0.0	0.0	0.3	2.3	13.5
10-19	0.0	0.0	0.0	0.0	0.0	0.1	0.6	3.1	8.7
20-29	0.0	0.0	0.0	0.0	0.0	0.1	0.6	2.3	2.4
30-39	0.0	0.0	0.0	0.0	0.0	0.0	0.7	1.8	3.2
40-49	0.0	0.0	0.0	0.0	0.1	0.3	1.1	2.2	0.0
50-59	0.0	0.0	0.0	0.1	0.4	1.2	3.6	5.0	3.2
60-69	0.0	0.0	0.1	0.5	2.0	5.6	8.8	12.2	10.3
70-79	0.0	0.0	0.5	3.0	6.8	12.4	17.2	20.5	24.6
80-89	0.0	0.0	2.3	4.6	9.8	13.1	15.9	16.3	9.5
90-99	0.0	33.4	33.0	31.0	28.5	24.8	19.7	12.4	7.9
100	100.0	66.6	64.2	60.8	52.5	42.5	31.5	21.9	16.7
Cohort Surviving (per 1,000)	999	988	969	929	827	625	306	84	15
Average PHE per Life Year Level	1.000	0.983	0.978	0.967	0.944	0.908	0.850	0.750	0.621

Source: POHEM simulation run, sample size 5,000.

G. Where to Obtain Data

POHEM has the potential to use far more data than the statistical systems of any one country can currently support. This poses problems of comparability if results from different countries use the same or similar methodology but data of widely varying quality and detail. There are, however, a number of ameliorating considerations. First, the POHEM methodology is "upward compatible" with lifetable methods. Any lifetable style model relies for input on a set of transition probabilities, and these exact same transition probabilities can be used as inputs to POHEM to produce identical results (except for monte carlo sampling error).⁹ Thus, given the initial investment in the simulation software infrastructure, POHEM style results should never be more difficult to produce or less internationally or intertemporally comparable than current life expectancy or emerging disability-free life expectancy estimates.

⁹ Indeed, such comparisons should be part of the validation process for microsimulation models like POHEM.

In fact, the easiest way to reduce the data requirements of POHEM is to use it in a "stripped down" version with a reduced range of inputs. For example, simple age-specific marriage and divorce rates can be used rather than multivariate and longitudinally based hazard functions available for the union formation and dissolution modules.

Second, there is a substantial degree of similarity between humans in different countries. At our current rough levels of quantitative knowledge of disease onset and progression, it is not necessarily unreasonable to assume identical biological disease processes in different countries (though differing distributions of risk factors should probably be taken into account). This affords the prospect of collaborative efforts that pool the fragmentary epidemiological knowledge from a range of countries. Also, it may turn out (based on cross-national research in the EuroQuol project¹⁰) that the V() function is broadly similar across many OECD countries. If this is the case, then another set of data that can be used by POHEM need not be produced by each country, but could be borrowed from another.

H. Next Steps

The POHEM framework can be used to produce a range of health statistics. One direction of work is to include explicit representations of transitions among states of functional limitations. This will permit a more direct calculation of extensions of disability-free life expectancy type measures. A second direction is to include explicit representation of a wider range of risk factors and diseases. A third is to begin to include linkages to the health care system. Finally, it is straightforward to increase the range of statistics and measures produced by POHEM. We shall discuss these in turn.

1. Functional Limitations

So far, POHEM incorporates four risk factors and four diseases. Health status values are assigned directly from diseases. However, it is more natural to think of these health status valuations being based on functional limitations (e.g. chest pain), which are in turn based on diseases (e.g. CHD).

Data to connect clinical disease descriptions with self-reported functional limitations are not generally available (though such data could be assembled). Thus, work is currently underway to add an autonomous block of transition probability functions for functional limitations. These transition probabilities are independent of diseases and risk factors, but do depend on age and sex. They are derived from straightforward cross-tabulations of the 1986 HALS (post-censal Health and Activity Limitations Survey) data. In effect, a "special case" simplified version of POHEM is being constructed to allow disability-free life expectancy-like calculations; the risk factor and disease modules will be "turned off".

These transition probabilities among levels of severity and kinds of functional limitation can be estimated directly from HALS to some extent by using the responses to associated questions on how long it has been since the current limitation began. In addition, the technique developed by Gentleman and Robertson (1989) for estimating transition probabilities for multivariate risk factor distributions from cross-sectional data can be applied to multiple dimensions of functional limitation, as well as institutionalization.

10 Personal communications with Prof. Alan Williams, York University, England.

PHEs can then be computed directly from functional limitations for each person-year in the explicitly simulated cohort sample of complete individual life paths. The health status values can be assigned as was done somewhat arbitrarily by Wilkins and Adams (1983), or more rigorously by using Torrance-style estimates.

This variant of POHEM makes much fuller use of available data, and is not subject to the kinds of limitations of disability-free life expectancy measures outlined earlier. In addition, the kinds of data that are being used for Canada exist in generally similar forms for a number of other countries. Thus, the POHEM approach is transferrable; it can provide a common methodological foundation for internationally comparable Population Health Expectancy estimates -- provided similar disability and demographic transition probabilities are used. Moreover, this stripped-down version of POHEM will be able to generate estimates of the distribution of population health expectancy.

2. Risk Factors and Diseases

So far the development of POHEM has been illustrative, so that only a small segment of the epidemiological literature has been drawn upon. One area where more systematic compilations of the quantitative literature have been undertaken is in the development of "health risk appraisal" packages. While the use of these packages in individual counselling sessions may have questionable aspects, the data underlying the better efforts represent major syntheses of recent quantitative epidemiological research. Two notable examples are Health and Welfare Canada's *Evalu*Life* (1989) and the U.S. "Healthier People" (Carter Center, 1988).

These health risk appraisal packages provide a ready source of transition probabilities for an array of diseases (based on relative risks and mortality rates by cause of death) conditional on age, sex, and various risk factors. The data can be easily incorporated into POHEM and this is underway. The data can then be used in a statistically appropriate manner by applying them to (synthetic) representative population samples. This latter population-based rather than individual application is similar to the Dutch PREVENT package (Gunning-Schepers, 1988; Gunning-Schepers et al., 1989).

However, these data do not include morbidity. Thus, a significant extension of the concept of health risk appraisal is possible by including not just deaths from lung cancer, for example, but also the onset and progression of the disease, consequential impacts on functional limitations, and health state values. Reviews of the research literature could provide data on morbidity, i.e. disease onset and progression, in an analogous way to the collection of the original health risk appraisal mortality data (e.g. Spasoff, 1982).

3. Health Care

Due to the methods of financing the Canadian health care system such as global funding of hospitals, it is quite difficult to obtain average unit costs of many health care services and procedures -- for example, disaggregated by age, sex, and type of malady as in the lower array of Figure 1. However, some fragmentary data do exist, and good data should be forthcoming as the new hospital MIS (management info system) project comes to fruition.

POHEM can be straightforwardly extended to accommodate such health care "visits" and their associated unit costs. Subsequent disease and health status progression could then be adjusted to depend on such visits (not all of which need be assumed to result in improvements in health). For example, the CHD module based on Weinstein et al. (1987) already includes

Coronary Artery Bypass Graft surgery, though in a very simple way using U.S. data. More detailed Canadian data from a recent study will shortly be available and plans are underway to include them in POHEM.

With the addition of such data, the "costs of illness" in the sense of Cooper and Rice (1976) and Hartunian et al. (1981) could be estimated. (Recall that these include both direct health care costs and the indirect costs of foregone earnings. The latter can be drawn from the companion DEMOGEN results.)

4. A Family of Population Health Measures

There is clearly a wide range of health state data that can be incorporated into POHEM based on existing published research, further analysis of existing statistical data, and forthcoming administrative and sample survey data. POHEM provides an integrating framework that can draw these various kinds and sources of data together to produce higher level summary statistics and indices. This is in part analogous to the System of National Accounts where GDP represents an aggregation across many data feeder systems.

POHEM can also be used to create related families of summary indices. One analogy is the Consumer Price Index. There is not only the overall "all items" index, but also sub-indices for food and shelter, for example. Similarly, POHEM can be used to estimate PHEs for various age ranges, not unlike the current practice of estimating life expectancy at ages 20, 40, 65 and 75, as well as at birth. PHEs could also be estimated for various socio-demographic sub-groups -- for example never-married or low income individuals. Of course, in these latter cases, the quality of the estimates will depend very much on the underlying epidemiological data that have been incorporated into the transition probability functions.

Furthermore, POHEM can be used to estimate answers to hypothetical "what if" questions, just as macro-econometric models use analyses of the System of National Accounts data as the basis for simulating the impacts on Gross Domestic Product of various changes in macro-economic policy, or of external shocks. One example of a family of generally accepted "what if" statistics in the health area is life expectancies and Potential Years of Life Lost (PYLL) derived from cause-deleted life tables.

POHEM can be used to generate exactly analogous statistics. It is a straightforward matter to generate two POHEM samples, one representing the "base case", the other a variant where one cause of death has been eliminated (i.e. the vector of age-specific mortality rates from that cause is set to zero). Then PYLLs and cause-deleted life expectancies can be directly estimated by comparing the two samples. Moreover, this concept can be extended to risk factors such as smoking to get "risk factor-deleted" life expectancies and PHEs (e.g. Wolfson and Birkett, 1989).

Another application would be an extension of health risk appraisal to a population basis. Instead of using POHEM to generate a completely synthetic cohort, an actual sample of individuals could be taken as the starting point -- for example those in the 1985 Health Promotion Survey (Health and Welfare Canada, 1988). Then the completion of each individual's life path could be simulated by POHEM (say 25 times) given his or her current risk factors, and then once again assuming elimination of all "modifiable" risk factors.

This (computationally intensive) procedure would yield two figures from POHEM for each individual -- a life expectancy reflecting current risk behaviour, and an "achievable" life expectancy assuming "optimal" risk factor modification. POHEM could also generate PHEs in the same two cases that would take account of expected morbidity. From these figures, a set of Population Health Risk Indices could be readily calculated -- for example the average

difference between current PHE and achievable PHE. This average difference would give an indication for the scope of various health promotion and preventive initiatives. The inclusion of morbidity by using PHE rather than life expectancy might also suggest rather different priorities for health policy by showing the impact of chronic disease (e.g. Wigle, 1989).

I. Concluding Comments

This paper has described the POHEM methodology, and shown its relationship to a variety of summary statistics on health status, particularly generalizations of life expectancy such as disability-free life expectancy. POHEM represents a feasible approach to generating these kinds of statistics. Moreover, it can be readily extended to take account of a much wider set of quantitative disease onset and progression descriptions extant in the epidemiological literature, and to generate a broader and more sensitive range of summary health statistics. Finally, the structure and empirical foundations of POHEM are such that many countries should be able to use POHEM or similar microsimulation methodologies with their own national data.

References

- Adams, O.B. and D.N. Nagnur (1988), *Marriage, Divorce and Mortality: A life Table Analysis for Canada and Regions, 1980-1982*, Statistics Canada, Catalogue No. 84-536E, Ottawa, September.
- Bebbington, A.C. (1989), "Expectation of Life without Disability measured from the OPCS Disability Surveys", Discussion Paper 651, University of Kent at Canterbury, August.
- Berkman, L. and L. Breslow (1983), *Health and Ways of Living, The Alameda County Study*, Oxford University Press.
- Branch, L.G., "Active Life Expectancy in 10,000 Men and Women in Three Communities", a paper presented to the first workshop of the International Network on Health Expectancy, September 11-12, 1989, Quebec City.
- Brouard, N. and J.M. Robine (1989), "A Method of Calculation of Health Expectancy from Longitudinal Surveys on the Elderly People in France", a paper presented to the first workshop of the International Network on Health Expectancy, September 11-12, 1989, Quebec City.
- Carter Center of Emory University (1988), *Healthier People Health Risk Appraisal Program*, Atlanta, January.
- Caselli, G., J. Duchêne, V. Egidi, A. Santini, and G. Wunsch (1989), "A Matter of Life and Death -- Methodologies for the life history analysis of adult mortality", Paper prepared for the UN/WHO/CICRED Working Group on Socio-economic Differences in Mortality in the Industrialized Societies, Paris, December.
- Cooper, B.S. and D.P. Rice (1976), "The Economic Cost of Illness Revisited", *Social Security Bulletin*, Vol. 39, pp. 21-36.
- Crimmins, E.M., Y. Saito, and D. Ingegneri (1989), "Changes in Life Expectancy and Disability-Free Life Expectancy in the United States", *Population and Development Review*, Vol 15, No 2, June.
- Erickson, P., E.A. Kendall, J. Anderson, and R.M. Kaplan (1988), "Using Composite Health Status Measures to Assess the Nation's Health", National Center for Health Statistics, Mimeo.

- Forbes, W.F. and J. Barham (1989), "An Investigation of the Prevalence and Incidence of Dementia in Populations Aged 65 Years and Over", Research Report 2, Program in Gerontology, University of Waterloo, Waterloo, March.
- Gentleman, J.F. and D. Robertson (1989), "Smoothing Procedures for Simulated Longitudinal Microdata", Proceedings of the Statistics Canada Symposium on the Analysis of Data in Time, October, Ottawa.
- Gentleman, J.F., M.C. Wolfson, and B.P. Will (1990), "Microsimulation of Health Status and Health Costs in Canada: Radon as a Risk Factor for Lung Cancer", presented to the Second International Conference on Statistical Methods for the Environmental Sciences (Environmetrics), Como, Italy, 27-30 September
- Gunning-Schepers, L. (1988), *The Health Benefits of Prevention: A Simulation Approach*, Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit, Rotterdam.
- Gunning-Schepers, L.J., J.J. Barendregt and P.J. Van der Maas (1989), "Population Interventions Reassessed", *The Lancet*, March 4, pp 479-481.
- Hartunian, N.S., C.N. Smart, and M.S. Thompson (1981), *The Incidence and Economic Costs of Major Health Impairments*, Lexington Books.
- Hayes, B. (1989), "Alternative Methodologies for Life-Cycle Simulations of the Incidence of Taxation and Benefit Systems", Discussion Paper Number TIDI/137/October 1989, Taxation, Incentives and the Distribution of Income Programme, London School of Economics.
- Health and Welfare Canada (1988), *Canada's Health Promotion Survey, Technical Report*, Eds., I. Rootman, R. Warren, T. Stephens and L. Peters, Ottawa: Minister of Supply and Services Canada.
- Health and Welfare Canada (1989), *Evalu*Life Manual*, Health Promotion Directorate, Health Services and Health Promotion Branch, May.
- Hill, G. (1989), personal correspondence.
- Kaplan, R.M. and J.W. Bush (1982), "Health related quality of life measurement for evaluation research and policy analysis", *Health Psychology*, Vol. 1, pp. 61-80.
- Katz, S., L.G. Branch, and M.H. Branson (1983), "Active Life Expectancy", *New England Journal of Medicine*, 309, 1218-24.
- Loewy, J.W., A.S. Kapadia, B. Hsi, and B.R. Davis (1989), "Statistical Methods Which Distinguish Between Attributes of Assessment: Prolongation of Life vs. Quality of Life", paper presented to the Joint Statistical Meetings, American Statistical Association, Washington, D.C., August.
- Loomes, G. and L. McKenzie (1989), "The Use of QALYs in Health Care Decision Making", *Social Science and Medicine*, Vol. 28, No. 4, pp. 299-308.
- Manton, K. and E. Stallard (1990), "Cross-sectional Estimates of Active Life Expectancy for the U.S. Elderly and Oldest-old Populations", *Journal of Gerontology*.
- Marmot, M.G. (1986,) "Social Inequalities in Mortality: The social Environment" in R.G. Wilkinson (Ed.), *Class and Health: Research and Longitudinal Data*, London, Tavistock Press, 21-33.

- Nakamura, A. and M. Nakamura (1978), "On Microanalytic Simulation and its Application in Population Projection", *Journal of the Operational Research Society*, Vol.29, No.4, pp. 349-360.
- Mehrez, A. and A. Gafni (1989), "Quality-adjusted Life Years, Utility Theory, and Healthy-years Equivalents", *Medical Decision Making*, Vol. 9, pp. 142-142.
- Picot, G. (1989), "Modelling the Lifetime Employment Patterns of Canadians", *The Family in Crisis: A Population Crisis?* (ed. J. Legaré, T.R. Balakrishnan and R.P. Beaujot), Proceedings of a Colloquium organized by the Federation of Canadian Demographers and Sponsored by the Royal Society of Canada, University of Ottawa, November.
- Pommier, P. (1981), "Social Expenditure: Socialization of Expenditure? The French Experience with Satellite Accounts", *Review of Income and Wealth*, December.
- Robine, J.M. (1986), "Disability-Free Life Expectancy, General Indicators of the Health of Populations", Conseil des affaires sociales et de la famille, Scientific Report, Quebec.
- Rosser, R.M. and P. Kind (1978), "A scale of valuations of states of illness: Is there a social consensus", *International Journal of Epidemiology*, Vol. 7, pp. 247-358.
- Rowe, G. (1989), "Event History Analysis of Marriage and Divorce in Canada", *Proceedings of the Statistics Canada Symposium on the Analysis of Data in Time*, October, Ottawa.
- Rowe, G. and M.C. Wolfson (1990), "Biased Divorce: Validation of Marital Status Life Tables and Microsimulation Models", paper presented to the United Nations Economic Commission for Europe, Seminar on Demographic and Economic consequences and Implications of Changing Population Age Structure, Ottawa, Canada, 24-28 September.
- Spasoff, R.A., I.W. McDowell, P.A. Wright and G.C. Dunkley (1982), The Canadian Risk Factor Review Project. In: Proceedings of the 17th annual meeting of the Society of Prospective Medicine. Bethesda, Maryland: Society of Prospective Medicine, pp. 36-44.
- Statistics Canada (1988), *Health and Activity Limitation Survey Use's Guide*, Ottawa.
- Statistics Canada and the Ontario Ministry of Health (1989), *Ontario Health Survey Information Manual*, Ottawa.
- Teillet, P. (1988), "A Concept of Satellite Accounts in the Revised SNA", *Review of Income and Wealth*, December.
- Torrance, G.W. (1987), "Utility Approach to Measuring Health-Related Quality of Life", *Journal of Chronic Diseases*, Vol. 40, No. 6, pp. 5930-600.
- Vaupel, J.W. and A.I. Yashin (1986), "Targeting lifesaving: Demographic linkages between population structure and life expectancy", *European Journal of Population*, Vol. 2, pp. 335-360.
- Weinstein, M.C. et al. (1987), "Forecasting Coronary Heart Disease Incidence, Mortality, and Cost: The Coronary Heart Disease Policy Model", *American Journal of Public Health*, Vol. 77, No. 11, and personal communication.
- Wigle, D.T., T. Wong, R. Lane and Y. Mao (1989), "The Economic Burden of Chronic Diseases in Canada", paper presented at the First National Conference on Chronic Diseases in Canada -- Challenges and Opportunities, Toronto, April 20-21.
- Wilkins, R. and O. Adams (1983), *Healthfulness of Life*, Institute for Research in Public Policy, Montreal.

- Wolfson, M.C., G. Rowe, J. Gentleman, and M. Tomiak (1990), "Earnings and Death -- Effects over a Quarter Century", Analytical Studies Branch Research Paper, Statistics Canada, Ottawa.
- Wolfson, M.C. (1989a), "A System of Health Statistics -- Toward a New Conceptual Framework for Integrating Health Data", 21st General Conference, International Association for Research in Income and Wealth, Lahnstein, West Germany, August and *Review of Income and Wealth* forthcoming.
- Wolfson, M.C. (1989b), "Divorce, Homemaker Pensions, and Lifecycle Analysis", *Population Research and Policy Review*.
- Wolfson, M.C. and N. Birkett (1989), "POHEM and an exploration of CHD", talk presented to the first annual Canadian epidemiology conference, August, Ottawa.
- Wolfson, M.C. and P. Glynn (1990), "A Template for Health Information", computer animated/graphics presentation to the Conference of Deputy Ministers of Health, Winnipeg, November 29, 1990.

ANALYTICAL STUDIES BRANCH
RESEARCH PAPER SERIES

No.

1. *Behavioural Response in the Context of Socio-Economic Microanalytic Simulation*, **Lars Osberg**
2. *Unemployment and Training*, **Garnett Picot**
3. *Homemaker Pensions and Lifetime Redistribution*, **Michael Wolfson**
4. *Modelling the Lifetime Employment Patterns of Canadians*, **Garnett Picot**
5. *Job Loss and Labour Market Adjustment in the Canadian Economy*, **Garnett Picot and Ted Wannell**
6. *A System of Health Statistics: Toward a New Conceptual Framework for Integrating Health Data*, **Michael C. Wolfson**
7. *A Prototype Micro-Macro Link for the Canadian Household Sector*, **Hans J. Adler and Michael C. Wolfson**
8. *Notes on Corporate Concentration and Canada's Income Tax*, **Michael C. Wolfson**
9. *The Expanding Middle: Some Canadian Evidence on the Deskilling Debate*, **John Myles**
10. *The Rise of the Conglomerate Economy*, **Jorge Niosi**
11. *Energy Analysis of Canadian External Trade: 1971 and 1976*, **K.E. Hamilton**
12. *Net and Gross Rates of Land Concentration*, **Ray D. Bollman and Philip Ehrensaft**
13. *Cause-Deleted Life Tables for Canada (1921 to 1981): An Approach Towards Analyzing Epidemiologic Transition*, **Dhruva Nagnur and Michael Nagrodski**
14. *The Distribution of the Frequency of Occurrence of Nucleotide Subsequences, Based on Their Overlap Capability*, **Jane F. Gentleman and Ronald C. Mullin**
15. *Immigration and the Ethnolinguistic Character of Canada and Quebec*, **Réjean Lachapelle**
16. *Integration of Canadian Farm and Off-Farm Markets and the Off-Farm Work of Women, Men and Children*, **Ray D. Bollman and Pamela Smith**
17. *Wages and Jobs in the 1980s: Changing Youth Wages and the Declining Middle*, **J. Myles, G. Picot and T. Wannell**
18. *A Profile of Farmers with Computers*, **Ray D. Bollman**
19. *Mortality Risk Distributions: A Life Table Analysis*, **Geoff Rowe**
20. *Industrial Classification in the Canadian Census of Manufactures: Automated Verification Using Product Data*, **John S. Crysdale**

21. *Consumption, Income and Retirement*, **A.L. Robb and J.B. Burbridge**
22. *Job Turnover in Canada's Manufacturing Sector*, **John R. Baldwin and Paul K. Gorecki**
23. *Series on The Dynamics of the Competitive Process*, **John R. Baldwin and Paul K. Gorecki**
 - A. *Firm Entry and Exit Within the Canadian Manufacturing Sector.*
 - B. *Intra-Industry Mobility in the Canadian Manufacturing Sector.*
 - C. *Measuring Entry and Exit in Canadian Manufacturing: Methodology*
 - D. *The Contribution of the Competitive Process to Productivity Growth: The Role of Firm and Plant Turnover.*
 - E. *Mergers and the Competitive Process.*
 - F. *(in preparation)*
 - G. *Concentration Statistics as Predictors of the Intensity of Competition*
 - H. *The Relationship Between Mobility and Concentration for the Canadian Manufacturing Sector*
24. *Mainframe SAS Enhancements in Support of Exploratory Data Analysis*, **Richard Johnson and Jane F. Gentleman**
25. *Dimensions of Labour Market Change in Canada: Intersectoral Shifts, Job and Worker Turnover*, **John R. Baldwin and Paul K. Gorecki**
26. *The Persistent Gap: Exploring the Earnings Differential Between Recent Male and Female Postsecondary Graduates*, **Ted Wannell**
27. *Estimating Agricultural Soil Erosion Losses From Census of Agriculture Crop Coverage Data*, **Douglas F. Trant**
28. *Good Jobs/Bad Jobs and the Declining Middle: 1967-1986*, **Garnett Picot, John Myles, Ted Wannell**
29. *Longitudinal Career Data for Selected Cohorts of Men and Women in the Public Service, 1978-1987*, **Garnett Picot and Ted Wannell**
30. *Earnings and Death - Effects Over a Quarter Century*, **Michael Wolfson, Geoff Rowe, Jane F. Gentleman and Monica Tomiak**
31. *Firm Response to Price Uncertainty: Tripartite Stabilization and the Western Canadian Cattle Industry*, **Theodore M. Horbulyk**
32. *Smoothing Procedures for Simulated Longitudinal Microdata*, **Jane F. Gentleman, Dale Robertson and Monica Tomiak**
33. *Patterns of Canadian Foreign Direct Investment Abroad*, **Paul K. Gorecki**
34. *POHEM - A New Approach to the Estimation of Health Status Adjusted Life Expectancy*, **Michael C. Wolfson**

For further information, contact the Chairperson, Publications Review Committee, Analytical Studies Branch, R.H. Coats Bldg., 24th Floor, Statistics Canada, Tunney's Pasture, Ottawa, Ontario K1A 0T6, (613) 951-8213.

